Digitale Rechtssubjekte? Haftung für das Handeln autonomer Softwareagenten

Gunther Teubner 2019-09-30T09:00:03

Wenn Softwareagenten autonome Entscheidungen treffen, bedeutet dies einen massiven Kontrollverlust menschlicher Akteure. Unausweichlich entstehen gravierende Verantwortungslücken, die das ausschließlich an Menschen orientierte geltende Recht nicht ausfüllen kann. Die herrschende Lehre im Zivilrecht jedoch hält das traditionelle Instrumentarium des Rechts für völlig ausreichend. Dem ist mit der folgenden These zu widersprechen: Ein eigener Rechtsstatus für Softwareagenten ist nötig, um die Gefahr einer ständig wachsenden Verantwortungslücke abzuwenden. Der Status müsste aber in einer funktionalen Sicht genau auf ihre Rolle digitaler Assistenz in Mensch-Maschinen-Interaktionen abgestimmt sein. Dafür dürfte nicht die volle Rechtsfähigkeit notwendig sein. Wohl aber sollte den Softwareagenten partielle Rechtssubjektivität zugeschrieben werden.

Drei neue Verantwortungsrisiken hat die Digitalität aufgeworfen: (1) das Autonomierisiko, das in eigenständigen Entscheidungen der Softwareagenten seinen Ursprung hat, (2) das Verbundrisiko, das auf die enge Kooperation von Mensch und Softwareagent zurückzuführen ist, und (3) das Vernetzungsrisiko, das entsteht, wenn Computer nicht isoliert agieren, sondern in enger Verflechtung mit anderen Computern. Wie soll das Recht darauf reagieren?

I. Autonomierisiko

Besonders einschneidend ist das Autonomierisiko, das vom prinzipiell unvorhersehbaren Verhalten selbstlernender Algorithmen erzeugt wird. Es verwirklicht sich dann, wenn Softwareagenten tatsächlich, wie es in den beteiligten Wissenschaften, besonders der Informationsphilosophie, der Aktor-Netzwerktheorie und der Systemtheorie, vielfach vertreten wird, als handlungsfähige Akteure auftreten.

1. Softwareagenten als kommunikationsfähige Akteure?

Wenn man nicht falschen Anthropomorphisierungen aufsitzen will, ist es notwendig, die Parallele zur Handlungsfähigkeit anderer nicht-menschlicher Akteure, zu den als juristischen Personen verfassten formalen Organisationen, zu ziehen. Man muss sich hier von der vertrauten Vorstellung lösen, das soziale Substrat juristischer Personen sei eine Vielheit konkreter Menschen. Der Kollektivakteur, wie ihn Talcott Parsons, Niklas Luhmann und andere definieren,

ist nicht eine Gruppe von Individuen, sondern – eine Kette von Mitteilungen. Unter der doppelten Voraussetzung, dass erstens diese Kommunikationskette über sich selbst kommuniziert, also eine Selbstbeschreibung herstellt, und dass zweitens gerade dieser Selbstbeschreibung kommunikative Ereignisse als Handlungen zugeschrieben werden, entsteht die soziale Realität eines Kollektivakteurs, das Substrat der juristischen Person. In ganz paralleler Weise sind auch Softwareagenten nicht als Maschinen zu verstehen, sondern – als Informationsflüsse. Diesen wird in der gesellschaftlichen Praxis unter bestimmten Bedingungen soziale Identität und Handlungsfähigkeit zugeschrieben. In beiden Fällen geht es um das Gleiche: um die gesellschaftliche Zuschreibung von Handlungsfähigkeit an – Kommunikationsprozesse.

Doch können wir wirklich davon ausgehen, dass Algorithmen kommunizieren, wie wir das bei Menschen, aber auch bei Organisationen voraussetzen? Auch hier wieder in aller Kürze der Vorschlag der soziologischen Systemtheorie: Die Antwort hängt davon ab, ob es in der gesellschaftlichen Kommunikation gelingt, die „Beiträge“ der Softwareagenten als kommunikative Ereignisse im strengen Sinne, also als Einheit von Information, Mitteilung und Verstehen zu aktivieren. Sofern der kommunikative Prozess Mensch-Computer Ereignisse identifiziert, die als „Mitteilungen“ der Algorithmen, welche eine bestimmte „Information“ enthalten, zu „verstehen“ sind, wird sich in der Tat in der Interaktion der Menschen mit Algorithmen ein genuines Sozialsystem herausbilden. Die „Antworten“, die wir von Software-Agenten auf unsere Anfragen erhalten, erfüllen somit alles, was den Dreiklang von Information, Mitteilung und Verstehen, der für Kommunikation im strengen Sinne erforderlich ist, ertönen lässt. Deshalb entsteht auch in der durchaus asymmetrischen Interaktion von Menschen und Algorithmen ein genuines Sozialsystem.

Ob nun ein Softwareagent als autonomer Akteur zu qualifizieren ist, ist die juristisch entscheidende Frage. Jeder Sozialkontext schafft sich seine jeweils eigenen Kriterien der Personalität, die Wirtschaft nicht anders als die Politik, die Wissenschaft, die Moralphilosophie oder das Recht. In der interdisziplinären Diskussion werden nun ganz unterschiedliche Kriterien angeboten, ab welchem Schwellenwert einem Softwareagenten eine als gradualisiert verstandene Autonomie zuzusprechen ist, etwa Denkfähigkeit, Kommunikationsfähigkeit, rationales Handeln, Nicht-Prognostizierbarkeit ihrer Konditionalprogramme, eigene Zielverfolgung, Lernfähigkeit, Selbstbewusstsein, Künstliche Intelligenz, oder gar ein digitales Gewissen. Die Unterschiede erklären sich aus dem jeweiligen Erkenntnisinteresse der beteiligten Disziplinen.

2. Rechtliche Kriterien für autonomes Handeln

Das Recht wiederum muss aus eigenem Erkenntnisinteresse die rechtlich relevante Grenze zwischen reaktiven und autonomen Handeln selbst festlegen. Denkfähigkeit oder künstliche Intelligenz wird in der Debatte immer wieder, als das eigentliche Kriterium vorgeschlagen. Doch für ihre Handlungsfähigkeit geht es nicht um die Frage: Welche Art von ontologischen Eigenschaften – Intelligenz, Geist, Seele, reflexive Kapazitäten, Einfühlungsvermögen – muss ein Softwareagent besitzen, um

als Akteur gelten zu können? Nicht die inneren Eigenschaften der Agenten, sondern die gesellschaftlichen Interaktionen, insbesondere wirtschaftlichen Transaktionen, an der die laufenden Operationen des Algorithmus teilnehmen, konstituieren den Algorithmus als Person, als kommunikationsfähigen Akteur.

Andere vorgeschlagene Kriterien wiederum gehen über die juristischen Minimalvoraussetzungen für Autonomie deutlich hinaus. Dazu gehören rationales Handeln, Selbständerung oder selbsttätiges Lernen. Es dürfte aber, wenn Schäden ausgeglichen werden sollen, nicht angemessen sein, erst dann mit rechtlichen Haftungsmechanismen einzusetzen, wenn die Programmkorrekturen nur von ihnen selbst, nicht aber von Programmierern vorgenommen werden. Schon gar nicht darf für eine rechtlich relevante Digitalautonomie eine vollausgebildete Künstliche Intelligenz, Empathie, Gefühle, Leidensfähigkeit, Selbstbewusstsein oder gar ein digitales Gewissen verlangt werden.

Entscheidung unter Ungewissheit – dies dürfte das rechtsrelevante Autonomiekriterium sein. Wenn diese Aufgabe an Softwareagenten delegiert wird, dann ist das Recht genötigt, ihnen Handlungsfähigkeit zuzusprechen. Softwareagenten handeln im Rechtssinne dann autonom, wenn (1) ein Softwareagent so programmiert ist, dass er zwischen Alternativen zu entscheiden hat, wenn (2) er diese Entscheidung als Optimierung verschiedener Kriterien treffen muss und wenn (3) ein Programmierer das Verhalten des Softwareagenten weder nachträglich erklären noch für die Zukunft voraussagen kann, sondern nur noch ex post korrigieren kann. Grund dafür ist der fundamentale Zusammenhang, der zwischen der Öffnung von Entscheidungsalternativen unter Ungewissheit, deren Delegation an nicht-menschliche Prozessoren und der dabei entstehenden Verantwortungsproblematik besteht.

II. Rechtsprobleme des Autonomierisikos

1. Digitale Verträge

Die geradezu revolutionäre Umwälzung, dass Menschen an Algorithmen die Aufgabe delegieren, selbständig Verträge abzuschließen und durchzuführen, berührt das Privatrecht in seinen Fundamenten. Denn selbstverständliche Voraussetzung war bisher, dass ausschließlich Menschen Willenserklärungen abgeben. Dennoch können viele Zivilrechtler in Computererklärungen keine fundamentalen Probleme erblicken, auch nicht im eigentlich kritischen Fall, wenn der Softwareagent nicht voll determiniert ist, sondern selbst autonom über den Vertrag entscheidet. Hier entzieht sich das konkrete Verhalten des Softwareagenten weitgehend der Kontrolle des Betreibers. Selbst vom Programmierer ist es nicht mehr im einzelnen determinierbar, nicht mehr prognostizierbar und nicht mehr kontrollierbar. Doch selbst für diese Situation eines Kontrollverlusts hält die herrschende Lehre daran fest, dass ausschließlich der menschliche Geschäftsherr selbst die Erklärung abgegeben hat.

Dies ist eine unhaltbare Fiktion. Die herrschende Lehre beschwört sogar Willensfreiheit und Menschenwürde, um jede Art von Rechtssubjektivität für

Softwareagenten als verfassungswidrig (!) zu blockieren. Welch ein Widerspruch! Wenn man es für verfassungswidrig hält, an Algorithmen Entscheidungen zu delegieren, dann müsste man diese Praxis verbieten, nicht aber dürfte man die Praxis klammheimlich zulassen und sie dann rechtlich so behandeln, als hätte keine Delegation stattgefunden.

Deshalb sollte man konsequent auf die Gegenposition übersetzen, wie es auch eine Reihe von Autoren vertreten: In genauer Entsprechung zu seiner realen Funktion in der Wirtschaftspraxis gibt der Softwareagent rechtlich eigenverantwortlich die Willenserklärung ab, handelt aber nicht im eigenen Namen, sondern im Namen des Prinzipals. Dafür braucht man ihnen keine volle Rechtsfähigkeit als juristische Person zuzuschreiben, sondern in funktionaler Sicht genügt die bloße Teilrechtsfähigkeit, die Stellvertretungsfähigkeit.

Mit Hilfe einer solchen Analogie lässt sich der wohl wichtigste Einwand ausräumen, dass den Softwareagenten die für den Vertragsschluss notwendigen Willenselemente, besonders das Erklärungsbewusstsein fehlen. Doch bekanntlich ist inzwischen das Erklärungsbewusstsein entpsychologisiert. Denn an die Stelle des subjektiven Erklärungsbewusstseins setzt der Bundesgerichtshof zwei objektive Normen: erstens, die soziale Norm, ob das konkrete Verhalten als eine bindende Willenserklärung verstanden werden darf, zweitens, die Pflicht des Erklärenden, diese soziale Norm zu erkennen und nicht gegen diese Pflicht fahrlässig zu verstoßen. Eine solche Kenntnis sozialer Normen kann durchaus in ein Softwareprogramm übersetzt werden.

2. Vertragliche Haftung: autonome Softwareagenten als Erfüllungsgehilfen?

Werden autonome Softwareagenten zur Erfüllung eines Vertrages eingesetzt und verstoßen sie gegen Vertragspflichten, dann gerät die herrschende Lehre erneut in große Schwierigkeiten. Da sie an der Prämisse unbeirrt festhält, dass selbst autonome Softwareagenten nicht handeln können und deshalb auch keine Erfüllungsgehilfen sein können, muss sie die Vertragsverletzung ausschließlich in der Person des Vertragsschließenden, also des Betreibers, suchen.Mit dieser fragwürdigen Konstruktion lässt die Lehre es zu, dass eine schwer erträgliche Haftungslücke entsteht. Kann der Betreiber nachweisen, dass der Betreiber selbst keine Vertragspflicht verletzt hat, dann ist dieser von jeglicher Haftung befreit. Der Kunde muss den vom Computer verursachten Schaden tragen. Und in Zukunft wird die Lücke sich ausweiten, je mehr Aufgaben der Vertragserfüllung an autonome Softwareagenten delegiert werden und je weniger der Betreiber das Verhalten des Softwareagenten vorhersehen oder beeinflussen kann.

Beide Schwierigkeiten lassen sich dagegen problemlos vermeiden, wenn man den autonomen Softwareagenten rechtlich zum Erfüllungsgehilfen erklärt. Da dies aber Rechtsfähigkeit voraussetzt, muss man ihm beschränkte Rechtssubjektivität zuschreiben. Gerade auch dann, wenn dem Geschäftsherrn selbst kein

Fehlverhalten vorzuwerfen ist, haftet der Geschäftsherr für jedes Verhalten des Softwareagenten, sofern dieses Vertragspflichten verletzt.

Letztlich ist es der Gleichbehandlungsgrundsatz, der hier die Haftung zwingend verlangt. Denn würde ein Mensch für die Vertragsdurchführung herangezogen, so haftete der Prinzipal nach § 278 BGB für dessen Pflichtverletzungen. Dann kann der Prinzipal aber nicht von der Haftung befreit sein, wenn er für die identische Aufgabe einen Softwareagenten heranzieht.

Subjektive Voraussetzungen sollten bei der Analogie zur Gehilfenhaftung noch weniger Schwierigkeiten machen als bei der Stellvertretung. Das Problem wird auch hier wieder durch Objektivierungstendenzen im heutigen Privatrecht entschärft.

3. Außervertragliche Haftung: Gefährdungshaftung als Königsweg?

Im Bereich außervertraglicher Haftung kritisieren die meisten Autoren, dass hier die technische Entwicklung eine erhebliche Haftungslücke aufgerissen hat, und fordern nachdrücklich das Eingreifen der Legislative. Der eigentliche Grund für das Versagen ist auch hier, dass sich die haftungsrechtlichen Normen ausschließlich darauf konzentrieren, ob Betreibern/Herstellern/Programmierern ein Fehlverhalten vorzuwerfen ist. Dagegen können nach geltendem Recht die Fehlentscheidungen, die Softwareagenten selbst trotz korrekten Verhaltens der menschlichen Beteiligten fällen, nicht sanktioniert werden.

Für die Zukunft scheint die Gefährdungshaftung der Königsweg zu sein, auf dem man dem digitalen Autonomierisiko erfolgreich begegnen könnte. Die meisten Autoren fordern mit großem Nachdruck eine entsprechende Gesetzgebung. Doch unterliegen sie einem fundamentalen Missverständnis. Die Leitprinzipien der Gefährdungshaftung können gar nicht als Vorbild dienen, da sie das digitale Autonomierisiko schlicht verfehlen. Das Risiko digitaler Entscheidungsautonomie ist prinzipiell andersgeartet als die Risiken, auf die es in den bisherigen Fällen der Gefährdungshaftung ankommt. Gefährdungshaftung setzt ein beim Einsatz gefährlicher Sachen, der aber wegen seines gesellschaftlichen Nutzens erlaubt, also rechtmäßig ist. Bei autonomen Softwareagenten kommt es aber gerade nicht auf die Sachgefahr eines falsch funktionierenden Computers, also auf das Kausalrisiko, an, sondern auf das Entscheidungsrisiko eines Computers, auf die ganz anders geartete Gefahr, dass sich dessen autonome Entscheidungen als Fehlentscheidungen herausstellen.

Bei der Gefährdungshaftung wird bekanntlich nicht nach Rechtswidrigkeit gefragt, bei Softwareagenten dagegen wird die Rechtswidrigkeit ihrer Entscheidung zum Dreh- und Angelpunkt der Haftung. Besonders bei den sogenannten „Rahmenrechten“, also etwa beim Persönlichkeitsrecht oder beim Recht am Unternehmen, aber auch bei sittenwidrigen oder wettbewerbswidrigen Handeln des Softwareagenten kann die Rechtswidrigkeit erst nach einer eingehenden Interessenabwägung festgelegt werden. Der zweite wichtige Mangel der Gefährdungshaftung betrifft den Umfang der

Haftung. Während Gefährdungshaftungsnormen regelmäßig die Haftung für Vermögensschäden ausschließen oder die Haftung auf nur wenige Rechtsgutsverletzungen beschränken, sind bei Softwareagenten solche Haftungsbeschränkungen nicht akzeptabel.

Die These heißt: Nicht die Gefährdungshaftung eines Betreibers für den rechtmäßigen Einsatz gefährlicher Anlagen, sondern nur Haftung des Betreibers für rechtswidriges Fehlverhalten des autonom entscheidenden Softwareagenten, dem partielle Rechtssubjektivität zuzuschreiben ist, dürfte als Grundprinzip einer strikten Haftung für digitales Handeln angemessen sein.

III. Verbundrisiko: Risiken des Mensch-Maschinen-Verbunds

Sehr viel unsicheres Gelände betritt man im Fall des Verbundrisikos, also, wenn der Verbund von Menschen und Algorithmen eigenständige Risiken erzeugt. Die Mensch-Maschinen-Interaktionen bilden Kollektivphänomene eigener Art. Nach der Netzwerk-Aktor-Theorie entsteht hier ein Ressourcenpool, der die beschränkten Handlungskapazitäten der Softwareagenten mit den kommunikativen Fertigkeiten realer Menschen kombiniert.

Ähnlich wie eine formale Organisation entwickelt der Mensch-Maschinen-Verbund selbst eine Innenperspektive, eine Selbstwahrnehmung, eine eigene Hierarchie von Präferenzen, eigene soziale Bedürfnisse und politische Interessen, die weder auf die Eigenschaften der beteiligten Menschen noch auf die der Algorithmen reduziert werden können. Die beteiligten Akteure handeln nicht für sich selbst, sondern „für“ den Hybriden als eine emergente Einheit, als Assoziation von Menschen und Nicht-Menschen. Den hier auftretenden Risiken, die aufgrund der fast nicht mehr auflösbaren Verflechtung der Einzelhandlungen von Menschen und Algorithmen entstehen, lässt sich besser begegnen, wenn man den Hybrid als solchen, als gemeinsamen Zurechnungspunkt für Handlungen, Rechte und Pflichten festlegt.

Sollte man die Rechtskonstruktion eines Mensch-Maschinen-Verbundes einführen? Dies ist eine bisher noch nicht ausprobierte Alternative: Die einzelnen rechtsgeschäftlichen Handlungen der Softwareagenten und die der beteiligten Menschen würden zu einer einheitlichen Handlung des Mensch-Maschinen-Verbundes zusammengezogen und würden sowohl Rechtsbindungen als auch Haftungsansprüche erzeugen.

IV. Vernetzungsrisiko: Risiken der Computer-Vernetzungen

Die Rechtstechnik der Personifizierung gerät jedoch dann an ihre Grenzen, wenn autonome Algorithmen in einem Multi-Algorithmen-System vernetzt sind. Personifizierung hat kein bestimmbares Substrat mehr bei komplexen Computervernetzungen. Das hier auftretende Vernetzungsrisiko zerstört Annahmen

über Individualität von Akteuren, die für die Zurechnung von Handlung und Verantwortung konstitutiv sind. Sowohl der Handlungsträger als auch die Kausalzusammenhänge sind dann schwierig, wenn nicht unmöglich, aufzuklären.

Ein Ausweg wäre, rechtliche Verantwortung nicht mehr „Personen“ als Handlungsträgern, sondern nur noch identifizierbaren Handlungen selbst, also einer „anonymen Matrix“ von sozialen und digitalen Prozessen selbst, zuzurechnen. Die eigentlichen Zurechnungspunkte für Verantwortung wären dann autonome Entscheidungen und nicht mehr die Entscheidungsträger. In einer solchen Situation bliebe nichts anderes übrig, als dass das Recht von sich aus autonom sorgfältig eingegrenzte Risiko-Pools konstruiert, um nicht zu sagen: dekretiert. Und sobald die Handlungen individueller oder kollektiver Akteure ebenso wie die Kalkulationen digitaler Akteure in einen solchen Raum geraten, würden sie alle als „Zwangsmitglieder“ eines solchen Risikopools haften – also nicht kraft privatautonomer Entscheidung, sondern kraft autoritativer Anordnung des staatlichen Rechts.

V. Ergebnis

Die drei neuen Formen eines digitalen Rechtsstatus für autonome Softwareagenten sind, je nachdem welches Risiko sich konkret verwirklicht, (1) Autonomierisiko: Akteur mit beschränkter Rechtssubjektivität, (2) Verbundrisiko: Mitglied eines Mensch-Maschinen-Verbunds, (3) Vernetzungsrisiko: Teilelement eines Risikopools. Ihre Konkretisierung ist darauf auszurichten, ob und wie die Verantwortungsdefizite bewältigt und Anreize zur Schadensprävention gesetzt werden können. Kernstück des digitalen Rechtsstatus aber ist, Algorithmen als handlungsfähigen Akteuren beschränkte Rechtssubjektivität zuzuerkennen.

Der Beitrag basiert auf: Teubner, Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten AcP 218 (2018), 155-205
